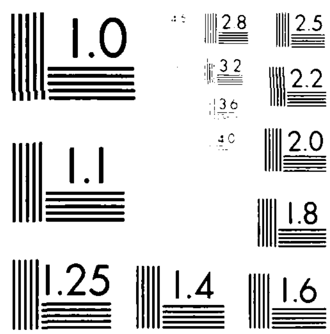
AD-A169 939 A NONPARAMETRIC QUANTILE ESTIMATOR: COMPUTATION(U) 1/1
SOUTH CAROLINA UNIV COLUMBIA DEPT OF STATISTICS
W J PADGETT MAY 86 TR-117 AFOSR-TR-86-0360
UNCLASSIFIED AFOSR-84-0156 F/G 12/1 NL




MICROCOPY RESOLUTION TEST CHART

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD-A169 939

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER Statistics Technical Report No. 117 | 2. GOVT ACCESSION NO. N/A | 3. RECIPIENT'S CATALOG NUMBER N/A |
| 4. TITLE (and Subtitle) A Nonparametric Quantile Estimator: Computation | | 5. TYPE OF REPORT & PERIOD COVERED Technical |
| | | 6. PERFORMING ORG. REPORT NUMBER AFOSR-TR- 86-0360 |
| 7. AUTHOR(s) W. J. Padgett | | 8. CONTRACT OR GRANT NUMBER(s) MIPR ARO 139-85 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of South Carolina Columbia, SC 29208 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS N/A |
| 11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709 | | 12. REPORT DATE May, 1986 |
| | | 13. NUMBER OF PAGES 15 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | | 15. SECURITY CLASS. (of this report) Unclassified |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC SELECTE JUL 2 4 1986 S D D

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Nonparametric quantile estimation; Right-censoring; Percentile interval; Bootstrap; Bandwidth selection.

DTIC FILE COPY

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A smooth nonparametric estimate of the quantile function from right-censored data is given by $Q_n(p) = h^{-1}\int_0^1 \hat{Q}_n(t)K((t-p)/h)dt$, $0 \le p \le 1$, where $K$ is a kernel function, $h$ is the "bandwidth," and $\hat{Q}_n$ is the product-limit quantile function. This report describes a computation procedure for data-based selection of a "best" bandwidth value to use in computing $Q_n(p)$ and for obtaining

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

# A NONPARAMETRIC QUANTILE ESTIMATOR: COMPUTATION*

by

W. J. Padgett

University of South Carolina
Statistics Technical Report No. 117
62G05-9

# DEPARTMENT OF STATISTICS

## The University of South Carolina
## Columbia, South Carolina 29208

A NONPARAMETRIC QUANTILE ESTIMATOR: COMPUTATION*

by

W. J. Padgett

University of South Carolina
Statistics Technical Report No. 117
62G05-9



May, 1986

Department of Statistics
University of South Carolina
Columbia, SC 29208



*Research supported by the U. S. Air Force Office of Scientific Research under grant number AFOSR-84-0156 and the U. S. Army Research Office under grant number MIPR ARO 139-85.

## ABSTRACT

A smooth nonparametric estimate of the quantile function from right-censored data is given by $Q_n(p) = h^{-1}\int_0^1 \hat{Q}_n(t)K((t-p)/h)dt$, $0 \leq p \leq 1$, where $K$ is a kernel function, $h$ is the "bandwidth," and $\hat{Q}_n$ is the product-limit quantile function. This report describes a computation procedure for data-based selection of a "best" bandwidth value to use in computing $Q_n(p)$ and for obtaining estimates of the bias and standard error as well as a nonparametric confidence interval for the true quantile value. The procedure is based on the bootstrap method for right-censored data. A listing of a FORTRAN program which performs the necessary calculations is provided, and examples of the procedure are given.

Key Words: Nonparametric quantile estimation; Right-censoring; Percentile interval; Bootstrap; Bandwidth selection.

# 1. INTRODUCTION

Right-censored data arise very naturally in life testing and reliability studies. For such data, it is important to be able to obtain good nonparametric estimates of various characteristics of the unknown lifetime distribution. This report concerns the computational procedure for a kernel-type nonparametric estimator of the quantile function of the lifetime distribution from right-censored data. This estimator was suggested by Padgett (1986), extending the complete sample results of Yang (1985). The large sample properties of the estimator, such as asymptotic normality and mean square convergence, were studied by Lio, Padgett and Yu (1986) and by Lio and Padgett (1985).

In this report, a procedure for calculation of the kernel-type quantile estimate from right-censored data is described, and a listing of a computer program in FORTRAN code is provided. In the procedure, the bootstrap is used to determine the approximate "optimal" bandwidth values to use for the kernel quantile estimates ("optimal" in the sense of smallest estimated mean squared error). The bootstrap method also provides estimates of the bias, mean squared error,and standard deviation of the estimator. In addition, a bootstrap confidence interval for the quantile, that is, a percentile interval, is obtained (see Efron, 1982, or Efron and Tibshirani, 1986, p. 68). See Padgett and Thombs (1986) for details on the performance of this computational procedure.

In section 2, the estimator is given. The computation procedure using the bootstrap method is described in section 3. Section 4 contains the listing of the FORTRAN source code and several examples on the utilization of the computer program.

## 2. THE NONPARAMETRIC QUANTILE ESTIMATOR

Let $X_1^o,\ldots,X_n^o$ denote the true survival times of $n$ items or individuals that are censored on the right by a sequence $U_1,\ldots,U_n$, which in general may be either constants or random variables. The $X_i^o$'s are nonnegative, independent, identically distributed random variables with common unknown distribution function $F_o$ and unknown quantile function $Q^o(p) \equiv F_o^{-1}(p) = \inf\{t\colon F_o(t) \geq p\}$, $0 \leq p \leq 1$.

The observed right-censored data are denoted by the pairs $(X_i,\Delta_i)$, $i=1,\ldots,n$, where

$$X_i = \min\{X_i^o,U_i\}, \qquad \Delta_i = \begin{cases} 1 & \text{if } X_i^o \leq U_i \\ 0 & \text{if } X_i^o > U_i. \end{cases}$$

Thus, it is known which observations are times of failure or death and which ones are censored or loss times. The nature of the censoring depends on the $U_i$'s. (i) If $U_1,\ldots,U_n$ are fixed constants, the observations are time-truncated. If all $U_i$'s are equal to the same constant, then the case of Type I censoring results. (ii) If all $U_i = X_{(r)}^o$, the rth order statistic of $X_1^o,\ldots,X_n^o$, then the situation is that of Type II censoring. (iii) If $U_1,\ldots,U_n$ constitute a random sample from a distribution $H$ (usually unknown) and are independent of $X_1^o,\ldots,X_n^o$, then $(X_i,\Delta_i)$, $i=1,2,\ldots,n$, is called a randomly right-censored sample. The observed value of $(X_i,\Delta_i)$ is denoted by $(x_i,\delta_i)$.

A popular estimator of the survival function $1-F_o(t)$ from the censored sample $(X_i,\Delta_i)$, $i=1,\ldots,n$, is the product-limit estimator of Kaplan and Meier (1958). The product-limit estimator, which was shown to be "self-consistent" by Efron (1967), is defined as follows. Let $(Z_i,\Lambda_i)$, $i=1,\ldots,n$, denote the

ordered $X_i$'s along with their corresponding $\Delta_i$'s. Then the product-limit estimator of $1-F_o(t)$ is

$$\hat{P}_n(t) = \begin{cases} 1, & 0 \le t \le Z_1, \\ \prod_{i=1}^{k-1} \left(\frac{n-i}{n-i+1}\right)^{\Lambda_i}, & Z_{k-1} < t \le Z_k,\ k=2,\ldots,n \\ 0, & Z_n < t. \end{cases}$$

The product-limit estimator of $F_o(t)$ is denoted by $\hat{F}_n(t) = 1-\hat{P}_n(t)$, and the size of the jump of $\hat{P}_n$ (or $\hat{F}_n$) at $Z_j$ is denoted by $s_j$. Note that $s_j = 0$ if and only if $Z_j$ is censored for $j < n$, i.e. if and only if $\Lambda_j = 0$. Define $S_i = \sum_{j=1}^{i} s_j = \hat{F}_n(Z_{i+1})$, $i=1,\ldots,n-1$, and $S_n \equiv 1$.

A natural estimator of $Q^o(p)$ is the product-limit (PL) quantile function $\hat{Q}_n(p) = \inf\{t: \hat{F}_n(t) \ge p\}$ (see Sander, 1975, or Cheng, 1984, for example). Since $\hat{Q}_n$ is a step function with jumps at the uncensored observations, it is desirable to obtain a smooth estimator of $Q^o$. One such estimator is the kernel estimator defined as follows (see Padgett, 1986): Let $\{h \equiv h_n\}$ be a "bandwidth" sequence of positive numbers so that $h_n \to 0$ as $n \to \infty$, and let $K$ be a bounded probability density function which is zero outside a finite interval and which is symmetric about zero. Then the kernel quantile function estimator is given by

$$\begin{aligned} Q_n(p) &= h^{-1} \int_0^1 \hat{Q}_n(t) K((t-p)/h)dt \\ &= h^{-1} \sum_{i=1}^{n} Z_i \int_{S_{i-1}}^{S_i} K((t-p)/h)dt,\ 0 \le p \le 1, \end{aligned} \tag{2.1}$$

where $S_0 \equiv 0$. Here, the kernel function $K$ is taken to be the triangular kernel on $[-1,1]$, i.e.

$$K(x) = \begin{cases} 1-|x|, & -1 \le x \le 1, \\ 0, & \text{otherwise}. \end{cases}$$

This function satisfies all conditions assumed by Padgett (1986), Lio, Padgett and Yu (1986), and Lio and Padgett (1985).

Procedures for choosing an "optimal" value of the bandwidth to use in calculating $Q_n(p)$ from a given right-censored data set and for estimating the standard error and obtaining an approximate confidence interval are described in the next section. These computations are based on the bootstrap for right-censored data, and a FORTRAN program for performing them is listed in section 4.

## 3. COMPUTATION PROCEDURES

The effective performance of the kernel estimator critically depends on the choice of the value of the bandwidth $h$, which is a "smoothing parameter." If $h$ is too small, not enough smoothing is done, and the estimate will be "rough," showing features which do not represent the true quantile function. If too much smoothing is done, i.e. $h$ is too large, important features of the true function may not be evident. Therefore, for a given set of right-censored data, a method of selecting a reasonable value of $h$ to use in calculating $Q_n(p)$ at each desired $p$ is needed. Ideally, the value of $h$ (which will depend on $p$ in general) that minimizes the mean squared error of $Q_n(p)$ (or some other criterion) should be used in computing $Q_n(p)$. However, that value of $h$ is not known and a method of estimating it from the censored data is required. Here, the bootstrap method of estimating the mean squared error of $Q_n(p)$ as a function of $h$ is employed. Then the value, $\hat{h}(p)$, which minimizes the bootstrap estimated mean squared error is selected as the "optimal" bandwidth for calculating $Q_n(p)$. The estimate of the mean squared error as a function of $h$ is obtained from 300 bootstrap samples taken from the given right-censored data set $(x_i,\delta_i)$, $i=1,\ldots,n$. This "data-based"

method of choosing the bandwidth value was mentioned by Padgett (1986) and was further investigated by Padgett and Thombs (1986).

The estimate of $Q^o(p)$ given by $Q_n(p)$ in equation (2.1) is calculated using the estimated bandwidth value, $\hat{h}(p)$, obtained from the procedure above. To estimate the bias, standard error, and a nonparametric confidence interval (i.e., a percentile interval), 1000 bootstrap samples are generated from the given right-censored data. For the $j$th bootstrap sample, an estimate $Q^b_{nj}(p)$ is obtained from formula (2.1), j=1,2,...,1000. The bias of $Q_n(p)$ is then estimated by

$$\text{BIAS} = \frac{1}{1000}\left[\sum_{j=1}^{1000} Q^b_{nj}(p) - \hat{Q}_n(p)\right],$$

and the standard error of $Q_n(p)$ is estimated from

$$\text{SE} = \left\{ \frac{1}{999}\left[ \sum_{j=1}^{1000} (Q^b_{nj}(p))^2 - \frac{\left(\sum_{j=1}^{1000} Q^b_{nj}(p)\right)^2}{1000} \right] \right\}^{1/2}.$$

Also, an estimate of the mean squared error of $Q_n(p)$ is obtained from $\text{MSE} = \text{SE}^2 + \text{BIAS}^2$. To calculate a nonparametric confidence interval for $Q^o(p)$, the central 95-percentile interval is obtained by selecting the 25th value of the ordered bootstrap estimates $Q^b_{nj}(p)$ as the lower bound and the 975th value in this ordering as the upper bound (see Efron, 1982, or Efron and Tibshirani, 1986, for a complete discussion of this type of interval).

## 4. THE COMPUTER PROGRAM AND EXAMPLES

A source code listing of the FORTRAN program which performs the computations described in section 3 is given in Table 1. That is, this program, chooses, based on the minimum bootstrap mean squared error criterion, the bandwidth value to use for calculation of $Q_n(p)$, calculates $Q_n(p)$, calcu-

lates the bootstrap estimates of the bias, mean squared error, and standard error, and gives the approximate 95-percentile interval for the true quantile. The IMSL (1982) subroutine GGUBS for generating uniform random numbers between zero and one is called by the FORTRAN program and must be available for use.

Table 2 gives an example of input data for the program in Table 1. The first line is the number of observations, $n = 15$. Lines 2-16 give the observed $(x_i, \delta_i)$ pairs, separated by spaces, line 17 gives the number $k$ of quantile values to be estimated $(1 \leq k \leq 10)$, and line 18 gives the $k=4$ values of 100p, separated by one or more spaces, at which $Q_n(p)$ is to be computed ($p = .05$, .10, .25, and .50 in this example). Table 3 shows the output for the data in Table 2. This output requires approximately one minute CPU time on a DEC VAX 785 computer. Tables 4 and 5 show the input and output for the mechanical switch data of Nair (1984), which has $n=40$ observations with 23 of them censored. Quantiles for $p=.05$ and $p=.25$ are calculated. These computations require about 3 minutes CPU time on a VAX 785.

Padgett and Thombs (1986) discuss the performance of this procedure of estimating quantiles from right-censored data. This nonparametric procedure seems to perform quite well, even for relatively small samples of size $n=20$ or so.

Table 1. FORTRAN Program Listing for Quantile Estimation

```
C  ESTIMATION OF QUANTILE USING KERNEL ESTIMATOR FROM RIGHT-
C      CENSORED DATA
C  WITH BOOTSTRAP EST. OF BIAS, VARIANCE, AND OPTIMAL BANDWIDTH
C  KERNEL - TRIANGULAR ON (-1,1)
       IMPLICIT REAL*8(A-H,O-Z)
        DIMENSION Z(300),IDEL(300),X(300),S(300),PF(300),ZS(300),
     1  IDS(300),W(300),P(10),QNPP(1000)
       REAL*4 W
       REAL*8 ISEED
       ISEED=22285.D0
       WRITE(6,100)
  100  FORMAT(' QUANTILE ESTIMATION FROM RIGHT-CENSORED DATA BY
     1 THE KERNEL METHOD ')
C  INPUT SAMPLE SIZE - INTEGER < 300
       READ *,N
C  INPUT DATA PAIRS Z,DELTA (SEPARATED BY SPACES)
       WRITE (6,148) N
  148  FORMAT('0SAMPLE SIZE =', I5/'0CENSORED SAMPLE:'/' ',9X,'X
     1   DELTA')
       DO 44 I=1,N
       READ *,Z(I),IDEL(I)
   44  WRITE (6,149) Z(I),IDEL(I)
  149  FORMAT(' ',F15.6,I4)
       XN=N
C  INPUT NUMBER OF PERCENTILES DESIRED - K
       READ *,K
C  INPUT PERCENTILES DESIRED - (UP TO 10 VALUES OF P, 0<P<100)
       READ *,(P(I),I=1,K)
       HNS=.01
       HNINC=.02
       HNMAX= .75
C  ORDER Z'S AND DELTAS
       NM1=N-1
       DO 15 I=1,NM1
        IP1=I+1
       DO 15 J=IP1,N
        IF(Z(I).LE.Z(J)) GO TO 15
        TEMP=Z(I)
        Z(I)=Z(J)
        Z(J)=TEMP
        ITEM=IDEL(I)
        IDEL(I)=IDEL(J)
        IDEL(J)=ITEM
   15  CONTINUE
       DO 900 III=1,K
        IFLAG=0
        NBSMAX=300
       PTP=P(III)/100.D0
       QMSESAV=1.D15
       HN=HNS
```

Table 1. Continued

```
    1  DO 17 I=1,N
       ZS(I)=Z(I)
   17  IDS(I)=IDEL(I)
       QBOOT=0.D0
       QBSQ=0.D0
       NBS=0
        I4=1
C   CALCULATE PL ESTIMATE AND S(I)=JUMP SIZES
    3  DO 20 I=1,NM1
        PF(I)=1.D0
       DO 20 J=1,I
        IF(IDS(J).EQ.0) GO TO 20
        XJ=J
        PF(I)=PF(I)*(XN-XJ)/(XN-XJ+1.D0)
   20  CONTINUE
       PF(N)=0.D0
       S(1)=1.D0-PF(1)
       DO 21 I=2,N
        IM1=I-1
   21  S(I)=PF(IM1)-PF(I)
       IF(NBS.GT.0) GO TO 45
C   CALCULATE PL QUANTILE ESTIMATE
       SJ=0.D0
       DO 35 J=1,N
        SJP=SJ+S(J)
        IF(S(J).EQ.0.D0) GO TO 35
        IF((PTP.LE.SJP).AND.(PTP.GT.SJ)) GO TO 36
        SJ=SJP
        GO TO 35
   36   QNHATP=ZS(J)
        GO TO 45
   35   CONTINUE
C   CALCULATE KERNEL QUANTILE ESTIMATE
   45  SJ=0.D0
       QP=0.D0
       DO 59 J=1,N
        A= (SJ-PTP)/HN
        A=DMAX1(A,-1.D0)
        SJ=SJ+S(J)
        XARG=(SJ-PTP)/HN
        B=DMIN1(XARG,1.D0)
        IF((A.GE.1.D0).OR.(B.LE.-1.D0)) GO TO 59
        IF(B.LE.0.D0) GO TO 51
        IF(A.GE.0.D0) GO TO 52
        QP=QP+ZS(J)*(B-A-(B*B+A*A)/2.D0)
        GO TO 59
   51   QP=QP+ZS(J)*(B-A+(B*B-A*A)/2.D0)
        GO TO 59
   52   QP=QP+ZS(J)*(B-A-(B*B-A*A)/2.D0)
   59  CONTINUE
```

Table 1. Continued

```
      IF(NBS.GT.0) GO TO 60
      QNP=QP
      GO TO 61
   60 QBOOT=QBOOT+QP
      QBSQ=QBSQ+QP*QP
       IF(IFLAG.EQ.0) GO TO 61
       QNPP(I4)=QP
       I4=I4+1
   61 NBS=NBS+1
      IF(NBS.GT.NBSMAX) GO TO 800
C  GENERATE N UNIFORM (0,1) RANDOM NUMBERS
C    USING IMSL SUBROUTINE GGUBS
      CALL GGUBS(ISEED,N,W)
C  GENERATE BOOTSTRAP SAMPLE
      DO 70 J=1,N
      WJ=DBLE(W(J))
      II=XN*WJ+1.D0
      ZS(J)=Z(II)
   70 IDS(J)=IDEL(II)
C ORDER ZS'S
      NM1=N-1
      DO 80 I=1,NM1
       IP1=I+1
      DO 80 J=IP1,N
       IF(ZS(I).EQ.ZS(J)) ZS(J)=ZS(J)+1.D-4
       IF(ZS(I).LT.ZS(J)) GO TO 80
       TEMP=ZS(I)
       ZS(I)=ZS(J)
       ZS(J)=TEMP
       ITEM=IDS(I)
       IDS(I)=IDS(J)
       IDS(J)=ITEM
   80 CONTINUE
      GO TO 3
  800 XNS=NBS-1
      QBIAS=QBOOT/XNS-QNHATP
      QMSE=(QBSQ-QBOOT**2/XNS)/(XNS-1.D0)+QBIAS**2
      IF(IFLAG.EQ.1) GO TO 860
      IF(QMSE.GE.QMSESAV) GO TO 850
      QMSESAV=QMSE
      HSAVE=HN
       IF(IFLAG.EQ.1) GO TO 860
  850 HN=HN+HNINC
      IF(HN.LT.HNMAX) GO TO 1
       IFLAG=1
       NBSMAX=1000
       HN=HSAVE
       GO TO 1
  860 QVAR=QMSE-QBIAS**2
```

Table 1. Continued

```
      WRITE(6,103)
  103 FORMAT('0  P        QN(P)       BIAS EST.      MSE EST.      VAR EST.
     1 BANDWIDTH USED')
      WRITE(6,151) PTP,QNP,QBIAS,QMSE,QVAR,HN
  151 FORMAT(' ',F4.2,4D13.5,F9.2)
      SE=DSQRT(QVAR)
      WRITE(6,152) SE
  152 FORMAT('0ESTIMATED STANDARD ERROR OF QN(P) =',D12.5)
       DO 751 I=1,999
       IP1=I+1
       DO 751 J=IP1,1000
       IF(QNPP(I).LE.QNPP(J)) GO TO 751
       TEMP=QNPP(I)
       QNPP(I)=QNPP(J)
       QNPP(J)=TEMP
  751  CONTINUE
  900  WRITE(6,153) QNPP(25),QNPP(975)
  153  FORMAT(' APPROXIMATE 95-PERCENTILE INTERVAL:',D13.5,
     1 ' TO ',D13.5)
      STOP
      END
```

Table 2. Example Input Data for Quantile Estimation Program

| | | |
|---|---|---|
| 15 | | n = sample size |
| 1.2837 | 0 | |
| .6636 | 0 | |
| .1827 | 0 | |
| 1.9805 | 0 | |
| .1393 | 0 | |
| .2796 | 1 | |
| .6807 | 1 | |
| .4247 | 1 | $x_i, \delta_i$ pairs |
| 1.1301 | 0 | |
| .3699 | 1 | |
| 1.9590 | 0 | |
| .1404 | 0 | |
| .1696 | 0 | |
| .1912 | 0 | |
| .4354 | 0 | |
| 4 | | k = number of quantiles to estimate |
| 5 10 25 50 | | k values 100p |

Table 3. Output for Data in Table 2

```
QUANTILE ESTIMATION FROM RIGHT-CENSORED DATA BY THE KERNEL METHOD
SAMPLE SIZE =   15
CENSORED SAMPLE:
         X        DELTA
       1.283700   0
       0.663600   0
       0.182700   0
       1.980500   0
       0.139300   0
       0.279600   1
       0.680700   1
       0.424700   1
       1.130100   0
       0.369900   1
       1.959000   0
       0.140400   0
       0.169600   0
       0.191200   0
       0.435400   0

  P       QN(P)       BIAS EST.     MSE EST.     VAR EST.    BANDWIDTH USED
0.05  0.25144D+00  0.27371D-01  0.27924D-01  0.27174D-01       0.11

ESTIMATED STANDARD ERROR OF QN(P) = 0.16485D+00
APPROXIMATE 95-PERCENTILE INTERVAL:  0.23801D+00 TO   0.57944D+00

  P       QN(P)       BIAS EST.     MSE EST.     VAR EST.    BANDWIDTH USED
0.10  0.28883D+00  0.10934D+00  0.58516D-01  0.46560D-01       0.29

ESTIMATED STANDARD ERROR OF QN(P) = 0.21578D+00
APPROXIMATE 95-PERCENTILE INTERVAL:  0.22141D+00 TO   0.10683D+01

  P       QN(P)       BIAS EST.     MSE EST.     VAR EST.    BANDWIDTH USED
0.25  0.77867D+00  0.37146D+00  0.20857D+00  0.70595D-01       0.73

ESTIMATED STANDARD ERROR OF QN(P) = 0.26570D+00
APPROXIMATE 95-PERCENTILE INTERVAL:  0.36779D+00 TO   0.13626D+01

  P       QN(P)       BIAS EST.     MSE EST.     VAR EST.    BANDWIDTH USED
0.50  0.14833D+01 -0.61632D+00  0.56409D+00  0.18424D+00       0.39

ESTIMATED STANDARD ERROR OF QN(P) = 0.42923D+00
APPROXIMATE 95-PERCENTILE INTERVAL:  0.56632D+00 TO   0.19792D+01
```

Table 4. Input Data for Mechanical Switch Failure Example

```
   40
1.151 0
1.170 0
1.248 0
1.331 0
1.381 0
1.499 1
1.508 0
1.534 0
1.577 0
1.584 0
1.667 1
1.695 1
1.710 1
1.955 0
1.965 1
2.012 0
2.051 0
2.076 0
2.109 1
2.116 0
2.119 0
2.135 1
2.197 1
2.199 0
2.227 1
2.250 0
2.254 1
2.261 0
2.349 0
2.369 1
2.547 1
2.548 1
2.738 0
2.794 1
2.883 0
2.884 0
2.910 1
3.015 1
3.017 1
3.793 0
 2
 5  25
```

Table 5. Output for Mechanical Switch Failure Data

QUANTILE ESTIMATION FROM RIGHT-CENSORED DATA BY THE KERNEL METHOD
SAMPLE SIZE = 40
CENSORED SAMPLE:

| X | DELTA |
|---|---|
| 1.151000 | 0 |
| 1.170000 | 0 |
| 1.248000 | 0 |
| 1.331000 | 0 |
| 1.381000 | 0 |
| 1.499000 | 1 |
| 1.508000 | 0 |
| 1.534000 | 0 |
| 1.577000 | 0 |
| 1.584000 | 0 |
| 1.667000 | 1 |
| 1.695000 | 1 |
| 1.710000 | 1 |
| 1.955000 | 0 |
| 1.965000 | 1 |
| 2.012000 | 0 |
| 2.051000 | 0 |
| 2.076000 | 0 |
| 2.109000 | 1 |
| 2.116000 | 0 |
| 2.119000 | 0 |
| 2.135000 | 1 |
| 2.197000 | 1 |
| 2.199000 | 0 |
| 2.227000 | 1 |
| 2.250000 | 0 |
| 2.254000 | 1 |
| 2.261000 | 0 |
| 2.349000 | 0 |
| 2.369000 | 1 |
| 2.547000 | 1 |
| 2.548000 | 1 |
| 2.738000 | 0 |
| 2.794000 | 1 |
| 2.883000 | 0 |
| 2.884000 | 0 |
| 2.910000 | 1 |
| 3.015000 | 1 |
| 3.017000 | 1 |
| 3.793000 | 0 |

| P | QN(P) | BIAS EST. | MSE EST. | VAR EST. | BANDWIDTH USED |
|---|---|---|---|---|---|
| 0.05 | 0.16482D+01 | 0.43077D-02 | 0.12651D-01 | 0.12632D-01 | 0.05 |

ESTIMATED STANDARD ERROR OF QN(P) = 0.11239D+00
APPROXIMATE 95-PERCENTILE INTERVAL: 0.14995D+01 TO 0.19955D+01

| P | QN(P) | BIAS EST. | MSE EST. | VAR EST. | BANDWIDTH USED |
|---|---|---|---|---|---|
| 0.25 | 0.21835D+01 | -0.11022D-01 | 0.18869D-01 | 0.18747D-01 | 0.03 |

ESTIMATED STANDARD ERROR OF QN(P) = 0.13692D+00
APPROXIMATE 95-PERCENTILE INTERVAL: 0.18969D+01 TO 0.25470D+01

## REFERENCES

Cheng, K. F. (1984), "On Almost Sure Representations for Quantiles of the Product-Limit Estimator with Applications," Sankhya, Ser.A, 46,426-443.

Efron, B. (1967), "The Two-Sample Problem with Censored Data," in Proceedings of the Fifth Berkeley Symposium (Vol. 4), Berkeley, CA: University of California Press, pp. 831-853.

Efron, B. (1982), "The Jackknife, the Bootstrap, and Other Resampling Plans," SIAM CBMS-NSF Monograph 38.

Efron, B. and Tibshirani, R. (1986), "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy," Statistical Science, 1, 54-75.

International Mathematical and Statistical Libraries, Inc. (1982), IMSL, Houston, Texas.

Kaplan, E. L. and Meier, P. (1958), "Nonparametric Estimation from Incomplete Observations," Journal of the American Statistical Association, 53, 457-481.

Lio, Y. L. and Padgett, W. J. (1985), "Some Convergence Results for Kernel-Type Quantile Estimators under Censoring," Statistics Tech. Rep. No. 108, Univ. of South Carolina.

Lio, Y. L., Padgett, W. J. and Yu, K. F. (1986), "On the Asymptotic Properties of a Kernel-Type Quantile Estimator from Censored Samples," Journal of Statistical Planning and Inference, to appear.

Nair, V. N. (1984), "Confidence Bands for Survival Functions with Censored Data: A Comparative Study," Technometrics, 26, 265-275.

Padgett, W. J. (1986), "A Kernel-Type Estimator of a Quantile Function from Right-Censored Data," Journal of the American Statistical Association, 81, 215-222.

Padgett, W. J. and Thombs, L. A. (1986), "Smooth Nonparametric Quantile Estimation Under Censoring: Simulations and Bootstrap Methods," Statistics Tech. Rep. No. 116, Univ. of South Carolina.

Sander, J. (1975), "The Weak Convergence of Quantiles of the Product-Limit Estimator," Technical Report No. 5, Stanford Univ., Dept. of Statistics.

Yang, S. S. (1985), "A Smooth Nonparametric Estimator of a Quantile Function," Journal of the American Statistical Association, 80, 1004-1011.

# REPORT DOCUMENTATION PAGE

| | |
|---|---|
| 1a. REPORT SECURITY CLASSIFICATION<br>UNCLASSIFIED | 1b. RESTRICTIVE MARKINGS |
| 2a. SECURITY CLASSIFICATION AUTHORITY<br>NA | 3. DISTRIBUTION / AVAILABILITY OF REPORT<br>Approved for Public Release; Distribution Unlimited |
| 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE<br>NA | |
| 4. PERFORMING ORGANIZATION REPORT NUMBER(S) | 5. MONITORING ORGANIZATION REPORT NUMBER(S) |
| 6a. NAME OF PERFORMING ORGANIZATION<br>Department of Statistics<br>6b. OFFICE SYMBOL *(If applicable)* | 7a. NAME OF MONITORING ORGANIZATION<br>AFOSR/NM |
| 6c. ADDRESS *(City, State, and ZIP Code)*<br>University of South Carolina<br>Columbia, SC 29208 | 7b. ADDRESS *(City, State, and ZIP Code)*<br>Bldg. 410<br>Bolling AFB, DC 20332-6448 |
| 8a. NAME OF FUNDING / SPONSORING ORGANIZATION<br>AFOSR<br>8b. OFFICE SYMBOL *(If applicable)*<br>NM | 9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER<br>AFOSR-84-0156 |
| 8c. ADDRESS *(City, State, and ZIP Code)*<br>Bldg. 410<br>Bolling, AFB, DC 20332-6448 | 10. SOURCE OF FUNDING NUMBERS<br>PROGRAM ELEMENT NO.: 6.1102F<br>PROJECT NO.: 2304<br>TASK NO.:<br>WORK UNIT ACCESSION NO.: |

11. TITLE *(Include Security Classification)*
A Nonparametric Quantile Estimator: Computation

12. PERSONAL AUTHOR(S)
W. J. Padgett

| 13a. TYPE OF REPORT | 13b. TIME COVERED | 14. DATE OF REPORT *(Year, Month, Day)* | 15. PAGE COUNT |
|---|---|---|---|
| Technical | FROM ______ TO ______ | May, 1986 | 15 |

16. SUPPLEMENTARY NOTATION

| 17. COSATI CODES | | |
|---|---|---|
| FIELD | GROUP | SUB-GROUP |
| | | |
| | | |

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Nonparametric quantile estimation; Right-censoring; Percentile interval; Bootstrap; Bandwidth selection.

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

A smooth nonparametric estimate of the quantile function from right-censored data is given by $\tilde{Q}_n(p) = h^{-1}\int_0^1 \hat{Q}_n(t)K((t-p)/h)dt$, $0 \le p \le 1$, where $K$ is a kernel function, $h$ is the "bandwidth," and $\hat{Q}_n$ is the product-limit quantile function. This report describes a computation procedure for data-based selection of a "best" bandwidth value to use in computing $\tilde{Q}_n(p)$ and for obtaining estimates of the bias and standard error as well as a nonparametric confidence interval for the true quantile value. The procedure is based on the bootstrap method for right-censored data. A listing of a FORTRAN program which performs the necessary calculations is provided, and examples of the procedure are given.

| | |
|---|---|
| 20. DISTRIBUTION / AVAILABILITY OF ABSTRACT<br>☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS | 21. ABSTRACT SECURITY CLASSIFICATION<br>UNCLASSIFIED |
| 22a. NAME OF RESPONSIBLE INDIVIDUAL<br>Maj. Brian W. Woodruff | 22b. TELEPHONE *(Include Area Code)*<br>(202) 767-5027<br>22c. OFFICE SYMBOL<br>NM |

**DD FORM 1473,** 84 MAR

83 APR edition may be used until exhausted
All other editions are obsolete

20. ABSTRACT (cont.)

estimates of the bias and standard error as well as a nonparametric confidence interval for the true quantile value. The procedure is based on the bootstrap method for right-censored data. A listing of a FORTRAN program which performs the necessary calculations is provided, and examples of the procedure are given.

END

DTIC

8-86